Weekly Report

1 Done

1.1 Revision

- Read the chapter.
- Removed typos.
- Made it easy to understand.
- Added latest visualization cases.
- Improved three exercises.

1.2 Paper Reading

- #### Manifold: A Model-agnostic Framework for Interpretation and Diagnosis of Machine Learning Models

Manifold is design to support interpretation, debugging, and comparison of machine learning models in an interactive manner.

Authors summarize design goals in both high-level and low level, which I think is quite interesting.

High-level phases	Low-level design goals
Inspection (Hypothesis)	1 Provide an overall summary of results generated by multiple models. 2 Enable an effective comparative analysis on model pairs: 2.1 Which model between has a higher accuracy among all test instances? 2.2 On which instances does one make a correct prediction but the other fail? 2.3 On which instances does them both make an agreement (both correct or both incorrect)? If both of them are correct, which model generates higher prediction scores (more confident)? 3 Enable an effective comparative analysis on a model and others: 3.1 On which instances does one make an agreement (disagreement) with the rest of the models? 3.2 Which models overall have a similar (different) behavior as one in terms of the model outcome?
Explanation (Reasoning)	1 Provide a visual summary for the feature distribution of the user-defined instance subset. 2 Enable an effective visual comparison of two different instance subsets regarding the feature distribution:

	2.1 Allow comparison of individual features. 2.2 Enable a quick Identification of the most discriminative features. 2.3 Provide a quantitative measure of the overall distribution similarity.
Refinement (Verification)	Generate feature (or model architecture) encoding strategies and iterate model refinement.

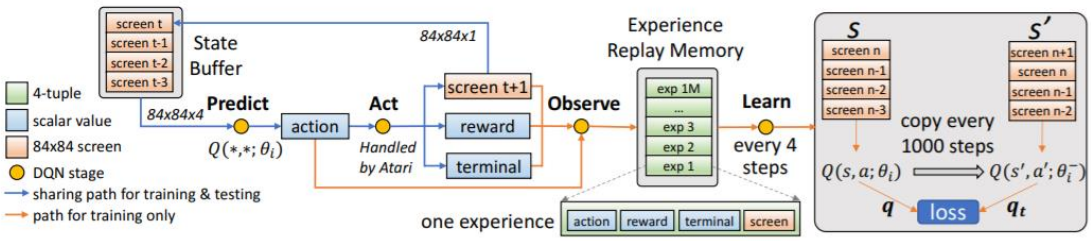
There are some concepts that I can mention in my new project.

- DQNViz: A Visual Analytics Approach to Understand Deep Q-Networks

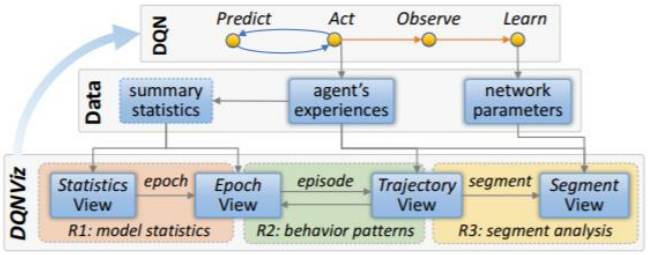
DQNViz explains Deep Q-Network (a deep reinforcement learning model) by a group of simple charts, like line charts and stacked area charts.

DQN aims to train an intelligent agent that can interact with an environment to achieve a desired goal. The model includes four major stages:

- Prediction: predicts for individual actions.
- Action: is handled by the environment.
- Observation: updates the experience replay memory.
- Learning: updates the Q-network by minimizing the loss at each iteration.



Combined with visualization, the framework is shown below.



The application scenario in the paper is the breakout game. And the question is how to maximize the total reward.

1.3 Others

- I applied for debit card and phone card.
- I sent documents to CSC.
- The University of California, Davis is closed for several days due to the bad air quality caused by fires. So, I just went to ViDi lab once in this week.

- I went to a birthday party and met some guys in ViDi.